PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : G06F 15/80	A1	(11) International Publication Number: WO 93/04438 (43) International Publication Date: 4 March 1993 (04.03.93)

(21) International Application Number: PCT/US92/06848

(22) International Filing Date: 13 August 1992 (13.08.92)

(30) Priority data:
746,038 16 August 1991 (16.08.91) US

(71) Applicant: THINKING MACHINES CORPORATION [US/US]; 245 First Street, Cambridge, MA 02142 (US).

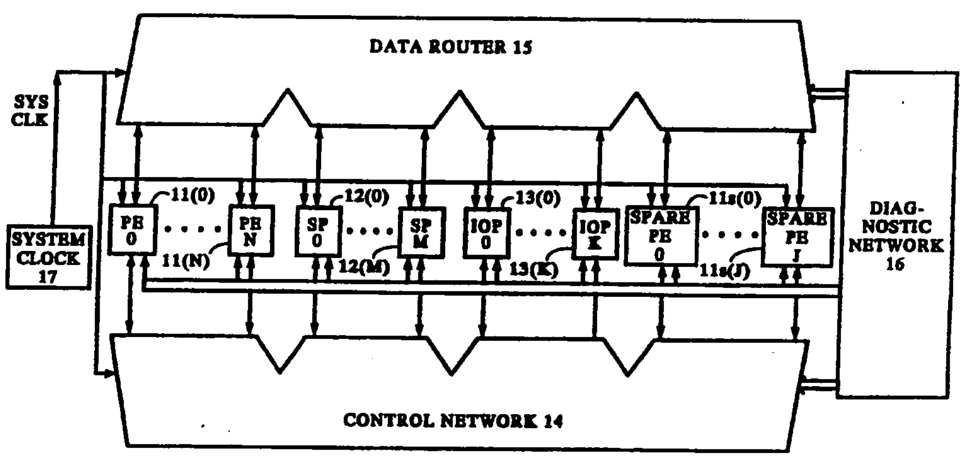
(72) Inventors: WELLS, David, S. ; 39 Bear Hill Road, Bolton, MA 01740 (US). ROWE, Eric, L. ; 140 Mill Street, Natick, MA 01760 (US). ISMAN, Marshall ; 11 Valley Spring Road, Newton, MA 02158 (US).

(74) Agent: JORDAN, Richard, A.; Thinking Machines Corporation, 245 First Street, Cambridge, MA 02142 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE).

Published
With international search report.

(54) Title: INPUT/OUTPUT ARRANGEMENT FOR MASSIVELY PARALLEL COMPUTER SYSTEM



SYSTEM 10

(57) Abstract

A computer comprising a plurality of processing elements (11) and an input/output processor (13) interconnected by a routing network (15). The routing network (15) transfers messages between the processing elements (11) and the input/output processor (13). The processing elements (11) perform processing operations in connection with data received from the input/output processor in messages transferred over the routing network and transferring processed data to the input/output processor in messages over the routing network, the processing elements being connected as a first selected series of leaf nodes. The input/output processor includes a plurality of input/output buffers connected as a second selected series of leaf nodes of the routing network for generating messages for transfer over the routing network to a series of processing elements forming at least a selected subset of the processing elements during an input/output operation.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	MN	Mongolia
AU	Australia	FR	France	MR	Mauritania
BB	Barbados	GA	Gabon	MW	Malawi
BE	Belgium	GB	United Kingdom	NL	Netherlands
BF	Burkina Faso	GN	Guinea	NO	Norway
BG	Bulgaria	GR	Greece	NZ	New Zealand
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	PT	Portugal
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	RU	Russian Federation
CG	Congo	KP	Democratic People's Republic of Korea	SD	Sudan
CH	Switzerland			SE	Sweden
CI	Côte d'Ivoire	KR	Republic of Korea	SK	Slovak Republic
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CS	Czechoslovakia	LK	Sri Lanka	SU	Soviet Union
CZ	Czech Republic	LU	Luxembourg	TD	Chad
DE	Germany	MC	Monaco	TG	Togo
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America

INPUT/OUTPUT ARRANGEMENT FOR MASSIVELY PARALLEL COMPUTER SYSTEM

BACKGROUND OF THE INVENTION

The invention relates generally to the field of digital computer systems, and more particularly to massively parallel computing systems.

A digital computer system generally comprises three basic elements, namely, a memory element, an input/output element and a processor element. The memory element stores information in addressable storage locations. This information includes data and instructions for processing the data. The processor element fetches information from the memory element, interprets the information as either an instruction or data, processes the data in accordance with the instructions, and returns the processed data to the memory element. The input/output element, under control of the processor element, also communicates with the memory element to transfer information, including instructions and the data to be processed, to the memory, and to obtain processed data from the memory.

Recently, computers have been developed which incorporate a large number of processing elements all of which may operate concurrently on generally the same instruction stream, but with each processing element processing a separate data stream. These processors have been termed "SIMD" processors, for "single-instruction/multiple-data" or, more generally "SPMD" processors, for "single-program/multiple-data" (collectively referred to herein as "SPMD")

SPMD processors are useful in a number of applications, such as image processing, signal processing, artificial intelligence, database operations, and computer simulation of a number of things, such as electronic circuits and fluid dynamics. In image processing, each processing element may be used to perform processing on a pixel ("picture element") of the image to enhance the overall image. In signal processing, the processors concurrently perform a number of the calculations required to perform such computations as the "Fast Fourier transform" of the data defining the signal. In artificial intelligence, the processors perform searches on extensive rule bases representing the stored knowledge of the particular application. Similarly, in database operations, the processors perform searches on the data in the database, and may also perform sorting and other operations. In computer simulation of, for example, electronic circuits, each processor may represent one part of the circuit, and the processor's iterative computations indicate the response of the part to signals from other parts of the circuit. Similarly, in simulating fluid dynamics, which can be useful in a number of applications such as weather predication and airplane design, each processor is associated with one point in space, and the calculations provide information about various factors such as fluid flow, temperature, pressure and so forth.

Typical SPMD systems include a SPMD array, which includes the array of processing elements and a router network, a control processor and an input/output component. The

input/output component, under control of the control processor, enables data to be transferred into the array for processing and receives processed data from the array for storage, display, and so forth. The control processor also controls the SPMD array, iteratively broadcasting instructions to the processing elements for execution in parallel. The router network enables the processing elements to communicate the results of a calculation to other processing elements for use in future calculations.

A deficiency in many types of computers having a number of processors, including SPMD computers, has been in the ability to quickly transfer data and other information between the input/output element and the processors.

SUMMARY OF THE INVENTION

The invention provides a parallel computer system including a new and improved input/output arrangement.

In brief summary, the invention in one aspect provides a computer comprising a plurality of processing elements and an input/output processor interconnected by a routing network. The routing network transfers messages between the processing elements and the input/output processor. The processing elements perform processing operations in connection with data received from the input/output processor in messages transferred over the routing network and transferring processed data to the input/output processor in messages over the routing network, the processing elements being connected as a first selected series of leaf nodes. The input/output processor includes a plurality of input/output buffers connected as a second selected series of leaf nodes of the routing network for generating messages for transfer over the routing network to a series of processing elements forming at least a selected subset of the processing elements during an input/output operation.

In another aspect, the invention provides an input/output processor including a plurality of input/output buffers connected to a series of leaf nodes of said routing network for generating messages for transfer over said routing network to a plurality of data receivers each connected to one of a second series of nodes of said routing network and identified by an address during an input/output operation. Each input/output buffer includes a transmit data buffer for buffering a plurality of data items each to be transmitted in a message to a data receiver in a message. A destination data receiver address and offset generator iteratively generates a destination data receiver address value and a destination offset value in response to the number of input/output buffers and the number of data receivers participating in the input/output operation.

BRIEF DESCRIPTION OF THE DRAWINGS

This invention is pointed out with particularity in the appended claims. The above and further advantages of this invention may be better understood by referring to the following description taken in conjunction with the accompanying drawings, in which:

Fig. 1 is a general block diagram of a massively parallel computer system constructed in accordance with the invention;

Figs. 2 is a diagram depicting the structure of message packets transmitted over the data router in the computer system depicted in Fig. 1;

Figs. 3A and 3B are functional block diagrams depicting the general structure of selected portions of the computer system of Fig. 1 useful in understanding the invention;

Figs. 4A and 4B are logic diagrams detailing the structure of circuits used in the portion depicted in Fig. 3A which generate information used in connection with generating portions of the message packets depicted in Fig. 2.

DETAILED DESCRIPTION OF AN ILLUSTRATIVE EMBODIMENT

Fig. 1 is a general block diagram of a massively parallel computer system 10 constructed in accordance with the invention. With reference to Fig. 1, system 10 includes a plurality of processing elements 11(0) through 11(N) (generally identified by reference numeral 11), scalar processors 12(0) through 12(M) (generally identified by reference numeral 12) and input/output processors 13(0) through 13(K) (generally identified by reference numeral 13). Input/output units (not shown), such as, for example, disk and tape storage units, video display devices, printers and so forth may be connected to the input/output processors to supply information, including data and program commands, for processing by the processing elements 11 and scalar processors 12 in the system, and may also receive processed data for storage, display and printing. The scalar processors 12 may also be connected to input/output units including, for example, video display terminals which permit one or more operators to generally control system 10. The system 10 may also include a plurality of spare processing elements 11s(0) through 11s(J) (generally identified by reference numeral 11s) which may be used as described below.

The system 10 further includes a control network 14, a data router 15 and a diagnostic network 16. The control network 14 permits one or more scalar processors 12 to broadcast program commands to processing elements 11. The processing elements 11 which receive the commands execute them generally concurrently. The control network 14 also permit the processing elements 11 to generate status information which they may supply to the scalar processors 12. The control network 14 is also used by the processing elements 11 to perform selected types of arithmetic operations, termed "scan" and "reduce" operations. The control network 14 may also be used to provide status and synchronization information among the processing elements 11.

The data router 15 transfers data among the processing elements 11, scalar processors 12 and input/output processors 13. In particular, under control of the scalar processors 12, the input/output processors 13 retrieve data to be processed from the input/output units and distributes it to the respective scalar processors 12 and processing elements 11. During processing, the scalar processors 12 and processing elements 11 can transfer data among themselves over the data router

15. In addition, the processing elements 11 and scalar processors 12 can transfer processed data to the input/output processors 13. Under control of the scalar processors 12, the input/output processors 13 can direct the processed data that they receive from the data router 15 to particular ones of the input/output units for storage, display, printing, or the like. The data router 15 in one particular embodiment is also used to transfer input/output commands from the scalar processors 12 to the input/output processors 13 and input/output status information from the input/output processors 13 to the scalar processors 12.

The diagnostic network 16, under control of a diagnostic processor (not shown in Fig. 1), facilitates testing of other portions of the system 10 to identify, locate and diagnose defects. The diagnostic processor may comprise one or more of the scalar processors 12. In addition, the diagnostic network 16 may be used to establish selected operating conditions in the other portions of the system 10.

The system 10 is synchronous, that is, all of its elements operate in accordance with a global SYS CLK system clock signal provided by a clock circuit 17.

One particular embodiment of system 10 may include hundreds or many thousands of processing elements 11 operating on a single problem in parallel under control of commands broadcast to them by the scalar processors 12. In that embodiment, the processing elements 11 operate in parallel on the same command on their individual sets of data, thereby forming a parallel computer system.

In addition, the system 10 may be dynamically logically partitioned, by logical partitioning of the control network 14, into multiple logical subsystems which may concurrently operate on separate problems or separate parts of a single problem. In that case, each partition includes at least one scalar processor 12 and a plurality of processing elements 11, the scalar processor 12 supplying the commands for processing by the processing elements in its partition. The spare processing elements 11s, which except for the positions of their connections to the control network 14 and data router 15 are otherwise similar to processing elements 11, may be used to substitute for failed processing elements 11 in a partition to augment the number of processing elements in a partition if there are insufficient processing elements 11 to form a partition with a desired number of processing elements 11, or to provide additional processing elements which may themselves be formed into partitions. In the following, unless otherwise stated explicitly, a reference to a processing element 11, in either the singular or plural, will also be taken as a corresponding singular or plural reference to a spare processing element 11s; that is, the processing elements 11 and spare processing elements 11s will be jointly referred to herein generally as processing elements 11.

Details of a control network 14, data router 15, and diagnostic network 16 used in one embodiment of the system 10 are described in International Application No. PCT/US91/07383, International Filing Date 3 October 1991, of Thinking Machines Corporation, entitled Parallel

Computer System (published under International Publication No. WO 92/06436 on 16 April 1992), and will not be repeated herein. In brief, both the control network 14 and data router 15 are generally tree-shaped networks (the data router 15 actually comprising a "fat tree") in which the processing elements 11, scalar processors 12 and input/output processors 13 are connected at the leaves. In addition, that International Application describes details of a network interface circuit included in the processing elements 11, scalar processors 12 and input/output processors 13 to enable them to communicate over the data router 15 and control network 14, which also will not be repeated herein.

The invention is generally directed to input/output operations in the system 10. Generally, input/output operations between an input/output processor 13 and processing elements 11 and the scalar processor 12 of a partition are controlled by the partition's scalar processor 12. The scalar processor 12 provides input/output command information to the processing elements 11 of its partition and the input/output processor(s) 13 to engage in the input/output operation. The scalar processor 12 provides the input/output command information to the input/output processor(s) 13 over the data router 15, to facilitate sharing of the input/output processors 13 among multiple partitions. In any case, the input/output command information provided by the scalar processor 12 to both the processing elements 11 and the input/output processor(s) 13 includes, inter alia, an operation identifier which identifies the input/output operation.

The aforementioned International Application further describes in detail the structure of message packets which the processing elements 11, scalar processors 12 and input/output processors 13 transmit over the data router 15 and control network 14 to effect information transfers thereamong. The invention described herein makes use of message packets transmitted over the data router 15, in particular input/output message packets having a particular structure which is depicted in Fig. 2. With reference to Fig. 2, an input/output message packet 2230 includes a message address portion 31, a message data portion 32 and a check portion 33. The message address portion is used to identify a path from the transmitting device to the intended recipient. The transmitting device and the intended recipient may be a processing element 11, a scalar processor 12 or an input/output processor 13. The message address portion 31 includes a HEADER portion, which contains a level identifier, and a series of down path identifiers DN "i" (index "i" is an integer from "M" to "1"). The level identifier in the HEADER portion identifies the lowest level in the tree that includes both the transmitting device and the intended recipient, and the data router 15 initially couples the input/output message packet 2230 from the transmitting device up to that level in the tree. Thereafter, the data router uses the successive down path identifiers DN "i" to steer the input/output message packet 2230 down the tree to the intended recipient.

The message data portion 32 includes a number of fields, including a message length field 34, a message tag field 35, a destination buffer identifier field 2231, a destination buffer offset field 2232 and a destination data field 2233. The message length field 34 identifies the length of the message

data portion 34. The message tag field 35 may contain operating system information identifying the packet as an input/output message packet 2230, from among other types of message packets which may be transmitted over the data router 15.

The contents of the destination buffer identification portion 2231 and the destination buffer offset portion 2232 provide information used by the receiving device, for example, a processing element 11(i) or scalar processor 12, in the case of input/output message packets 2230 transferred from an input/output processor 13, or by an input/output processor 13 in the case of an input/output message packet 2230 received thereby from a processing element 11(i) or a scalar processor 12. In particular, the contents of the destination buffer identification portion 2231 is derived from the input/output operation identifier, which is provided by the scalar processors 12 in their input/output commands. For example, if, as is typical, during input/output operations data is received by the receiver in an input/output buffer maintained thereby, the contents of the destination buffer identification portion 2231 may be used to identify the particular buffer into which the receiver may load the contents of the destination data portion 2233. The contents of the destination buffer offset portion 2232 identifies the particular location in the buffer into which the receiver is to load the contents of the destination data portion 2233. It will be appreciated that a number of distinct input/output operations may be performed in system 10 contemporaneously, with the input/output message packets 2230 having diverse values in their destination buffer identification portions 2231.

In addition, while the particular message transmitter, which may comprise either a processing element 11(i) or a scalar processor 12, on the one hand, or the input/output processor 13, on the other hand, may generate and transmit input/output message packets 2230 in the order in which they have the data to be transmitted, it will be appreciated that the message receivers may receive the input/output message packets 2230 in random order. The contents of the destination buffer offset portion 22 of each input/output message packet 2230 enables the receiver to properly order the data contained in the destination data portions 2233 of the received input/output message packets 2230 that are associated with the particular input/output operation as indicated by the contents of their destination buffer identification portions 2231.

Finally, the check portion 33 contains a cyclic redundancy check value which may be used to verify that the input/output message packet 2230 was correctly received.

The invention provides an arrangement for generating information for the message address portion 31 and destination buffer offset portion 2232 of an input/output message packet 2230.

A brief description of a parallel mode message transfer operation will be presented in connection with Figs. 3A and 3B. These Figs. schematically depict, respectively, a number of input/output buffer nodes 2201(0) through 2201(6) (Fig. 3A) comprising portions of an input/output processor 13 participating in an input/output operation with a partition of processing elements identified by reference numerals 11(0) through 11(5) (Fig. 3B). In particular, Fig. 3A schematically

represents, for each input/output buffer node 2201(i), a network interface 2277(i), a parallel send address/offset generator 2278(i) and a parallel mode buffer in the buffer memory 2223(i). Similarly, Fig. 3B schematically represents the network interface 202(i) and a memory buffer for each processing element 11(i). The network interfaces 2277(i) of sequentially-indexed input/output buffer nodes 2201(i) are connected as sequential leaves of the data router 15. Similarly, the network interfaces 202(i) of the sequentially-indexed processing elements 11(i) are connected as sequential leaves of the data router 15. The connections to data router 15 of the network interfaces 2277(i) of the input/output buffer nodes 2201(i) may be adjacent to the connections of the network interfaces 202(i) of the processing elements 11(i), or they may be separated. The number of input/output buffer nodes and processing elements participating in an input/output operation will be generally identified as "N" and "NPE," respectively.

As represented schematically in Fig. 3A, if, during the input/output operation, data is to be transferred in input/output message packets from an input/output device (not shown) to the processing elements, a device interface 2202 transfers data to the buffers of the input/output message buffers 2201(i) on a round-robin basis. That is, the device interface 2202 will transmit to input/output buffer node 2201(0) the first item of data, to input/output buffer node 2201(1) the second item of data, to input/output buffer node 2201(2) the third item of data, and so forth, where each "item of data" refers to the amount of data which it receives from the input/output device to be transmitted in an input/output message packet. After the device interface 2202 transmits an item of data to the last input/output buffer node to be participating in the input/output operation, here input/output buffer node 2201(7), it transmits the next item of data to input/output buffer node 2201(0), thereby ensuring that data is transmitted to the input/output buffer nodes in round-robin fashion.

The items of data transmitted to the input/output buffer nodes 2201(i) are arranged by the input/output device and device interface 2202 so that they will be directed to the processing elements 11(i) of increasing values of index "i," also on a round-robin basis with respect to the index of the processing element reference numeral. However, a selected number of sequential items of data directed to the input/output buffers 2201(i) may be intended for the same processing element, which number is termed herein a "striping factor," and which is generally identified as "C."

In addition, the items of data sequentially received by an input/output buffer node 2201(0) are stored at locations having successive offsets in the buffers of respective buffer memories 2223(i). In both Figs. 3A and 3B, the base of a buffer, that is, the location with a zero offset is depicted at the uppermost location in the respective buffer, and successive offsets are represented by the successively descending positions in the buffer.

Thus, for example, using the example depicted in Figs. 3A and 3B of seven input/output buffer nodes 2201(0) through 2201(6), six processing elements 11(0) through 11(5), and a striping factor of three, the data items for the first three messages for processing element 11(0) are

transferred from the device interface 2202 to input/output buffer nodes 2201(0) through 2201(2) and are represented in the respective buffer memories 2223(0) through 2223(2) as "PE(0) MSG(0)" through "PE(0) MSG(2)." The device interface 2202 next transmits the data items for the first three messages for processing element 11(1) to input/output buffer nodes 2201(3) through 2201(5) and are represented in the respective buffer memories 2223(3) through 2223(5) as "PE(1) MSG(0)" through "PE(1) MSG(2)." Thereafter, the device interface 2202 transmits the data item for the first message for processing element 11(2) to input/output buffer node 2201(6), and the data items for the second and third messages for the same processing element 11(2) to input/output buffer node 2201(0) and 2201(1). These data items are represented in the respective buffer memories 2223(6), 2223(0) and 2223(1) by the legends "PE(2) MSG(0)", "PE(2) MSG(1)" and "PE(2) MSG(3)," respectively. The device interface transmits the successive items of data to the input/output buffer nodes 2201 in the same way.

In the following, data items will be generally identified "PE(x) MSG(y)," where "x" identifies the processing element and "y" identifies the offset. With reference to Fig. 3A, it can be observed that the first data item PE(0) MSG(0) of the first series of data items provided by the input/output device to be transmitted to processing element 11(0) is in the buffer of buffer memory 2223(0) of input/output buffer node 2201(0) at offset zero. The last data item PE(5) MSG(2) of the first series of data items to be transmitted to the last processing element 11(5) is in the buffer of buffer memory 2223(3) of input/output buffer node 2201(3) at offset 2. This set of buffer locations across the buffers of the group of input/output buffer nodes 2201(0) through 2201(6) that are participating in an input/output operation will be termed a "frame."

More generally, a frame is a set of buffer locations, across the buffers of the input/output buffer nodes 2201(i) participating in an input/output operation, extending from the first data item PE(x) MSG(y) in a series to be transmitted as a stripe to the first processing element 11(0) to the last data item PE(x) MSG(y) in the corresponding series to be transmitted as the same stripe to the last processing element 11(5). Each of the sequence of frames in the buffer memories 2223(i) will be identified by a frame identifier value. That is, the frame containing locations from offset zero of the buffer of buffer memory 2223(0), which contains data item PE(0) MSG(0), to offset two of the buffer of buffer memory 2223(3), which contains data item PE(5) MSG(2), will be identified as frame zero. Similarly, the frame containing locations from offset two of the buffer of buffer memory 2223(4), which contains data item PE(0) MSG(3) to the offset of the buffer memory which contains data item PE(5) MSG(5) (not shown) will be identified as frame one, and so forth.

The series of data items PE(x) MSG(y) in a frame that are to be transferred to a particular processing element 11(i) or scalar processor 12 will be termed a "stripe." Each of the sequence of stripes in the buffer memories will be identified by a stripe offset value, which identifies the offset of the stripe from the beginning of a frame. That is, in the first frame, the data items in the first stripe, that is, the stripe at offset zero and containing data items PE(0) MSG(0) through PE(0) MSG(2), are

to be transferred to the first processing element 11(0) in the series participating in the input/output operation. Similarly, data items in the second stripe, that is, the stripe at offset one and containing data items PE(1) MSG(0) through PE(1) MSG(2) are to be transferred in input/output message packets 2230 to the second processing element 11(1) in the series participating in the input/output operation, and so forth.

Each buffer location in the frame will also be termed a slot and will be identified by a slot offset value identifying the offset of the particular slot from the beginning of a frame. Thus, the location of offset zero of the buffer of buffer memory 2223(0) has a slot offset value of zero, the location of offset zero of the buffer of buffer memory 2223(1) has a slot offset value of one, and so forth. The location of offset two of the buffer of buffer memory 2223(3), which contains data item PE(5) MSG(2), has a slot offset value of fourteen. Similarly, the location of offset two of the buffer of buffer memory 2223(4), which contains data item PE(0) MSG(3), which is the first slot of the second frame, has a slot offset value of zero. It will be appreciated that the number of slots, and thus the number of data items PE(x) MSG(y), in a frame, corresponds to the number of processing elements NPE times the striping factor C.

As also described above, the input/output buffer nodes 2201(i) transmit the successive data items PE(x) MSG(y) in their respective buffer memories to the processing elements 11(i), as represented on Fig. 3B. As shown on Fig. 3B, each processing element receives the messages containing the data items for its index "x" in the data item identification PE(x) MSG(y), and stores them in successive offsets "y." Thus, it will be appreciated that the indices "x" and "y" in the data item identification PE(x) MSG(y) reference the processing element identification and the offset, respectfully.

It will further be appreciated that complementary operations will occur in an input/output operation in the reverse direction to transfer data items from the successive buffer offsets of the processing elements 11(i), through the buffer memories 2223 of the input/output buffer nodes and to the input/output device. In that case, however, the processing element 11(0) will transmit the first three data items PE(0) MSG(0), PE(0) MSG(1), and PE(0) MSG(2) in its buffer to the input/output buffer nodes 2201(0) through 2201(2), and so forth. Thus, the input/output buffer node identifications used in the address portions 31 of the input/output message packets will be related to the index "y" of the data item identification PE(x) MSG(y), and the buffer offset will be related to the index "x."

The parallel send address/offset generator 2278(i) in each input/output buffer node 2201(i) generates, for each input/output message packet, information providing the processing element identification "x," in particular, the address of the processing element, which the network interface 2277(i) uses to generate the information for the message address portion 31 of the input/output message packet 2230. In addition, the parallel send address/offset generator 2278(i) generates the offset "y" for the data item PE(x) MSG(y). In this operation, the parallel send address/offset

generator 2278(i) operates using several items of information, including:

(a) the number of input/output buffer nodes "N" participating in the input/output operation,

(b) the striping factor "C,"

(c) the number of processing elements "NPE" participating in the input/output operation,

(d) the index "i" of the input/output buffer node 2201(i), and

(e) the address of the first processing element 11(0) in the partition participating in the input/output operation, relative to the base of the processing element of the system.

These items of information may be provided when initiating the input/output operation.

From these items of information, the parallel send address/offset generator 2278 may determine the following initial values used in connection with generating the first input/output message packet 2230 in the input/output operation:

(a) an initial destination processing element address value,

(b) an initial destination offset value, comprising (i) an initial offset base value and (ii) an initial offset delta value, both of which the parallel send address/offset generator 2278 will use to determine an initial destination processing element buffer offset value, and

(c) an initial slot value,

and the following incrementation values used in connection with generating subsequent input/output message packets 2230, if any, in the input/output operation:

(d) a destination processing element address incrementation value,

(e) offset incrementation values, including (i) an offset base incrementation value and (ii) an offset delta incrementation value, and

(f) a slot incrementation value.

It will be appreciated that these values may alternatively be provided when initiating the input/output operation.

A parallel send address/offset generator 2278(i), a detailed block diagram of which is depicted in Figs. 4A and 4B, includes four general sections, namely, a destination processing element address generating section 2310, an offset delta generating section 2311, an offset base generating section 2312, and a slot count section 2313. The offset base generating section 2312 and offset delta generating section 2311 generate, respectively, OFFSET BASE and OFFSET DELTA signals which are coupled to an adder 2314. The adder 2314, in turn, generates DEST OFFSET destination offset signals representing a value corresponding to the arithmetic sum of the values represented by the OFFSET BASE and OFFSET DELTA signals, which are latched in a latch 2315. The parallel send address/offset generator 2278(i) also couples the DEST OFFSET signals over bus 2287, to be used by

the network interface 2277 in generating the destination buffer offset portion of an input/output message packet.

A destination offset value is essentially formed from two components, one relating to the frame identifier of the frame containing the data item PE(x) MSG(y) being transmitted by the input/output buffer node, and the other relating to the offset of the slot containing the data item PE(x) MSG(y) in the series within the frame that is to be transferred to the same processing element 11(i) or scalar processor 12. In particular, the binary-encoded value represented by the OFFSET BASE signals, generated by the offset base generating section 2312, represents the component relating to the frame identifier value. Similarly, the binary-encoded value represented by the OFFSET DELTA signals represents the component relating to the position of the slot containing the data item in the sequence within a stripe.

The offset base generating section uses the initial offset base value and the offset base incrementation value, as well as the striping factor "C" and a BUMP OFFSET BASE signal from the slot count section 2313 in generating the OFFSET BASE signal. The initial offset base value for a parallel send address/offset generator 2278(i) relates to the frame of the first data item to be transmitted during the input/output operation. The frame identifier value of the data item PE(x) MSG(y) corresponds to the greatest integer in the quotient of (a) the input/output buffer node's index "i", divided by (b) the number of data items in a frame, which corresponds to the striping factor "C" times the number "NPE" of processing elements 11(i) participating in the input/output operation. The frame identifier, in turn, is multiplied by the striping factor "C," since for each subsequent frame the base offset value for the first data item PE(x) MSG(y) in each stripe corresponds to this value.

The offset base incrementation value is related to the number of frames that the input/output buffer node will increment between transmission of input/output message packets 2230. It will be appreciated that the number of frames will correspond to the greatest integer in the quotient of (a) the number "N" of input/output buffer nodes 2201(i) participating in the input/output operation, divided by (b) the number of slots in a frame, that is, the striping factor "C" times the number "NPE" of processing elements 11(i) participating in the input/output operation. This value is also multiplied by the striping factor "C," since the base for each subsequent frame will begin with a value corresponding to the frame identifier times the striping factor.

It will be appreciated that, if the number "N" of input/output buffer nodes 2201(i) participating in the input/output operation is not a multiple of the number of slots in a frame, the offset of the slot containing the data item PE(x) MSG(y) being transmitted will change for each subsequent input/output message packet. The change in the slot offset corresponds to the remainder of the quotient of (a) the number "N" of input/output buffer node 2201(i) participating in the input/output operation, divided by (b) the number of slots in a frame, that is, the striping factor "C" times the number "NPE" of processing elements 11(i) participating in the input/output operation, which remainder, in turn, corresponds to the number "N" modulo the number of slots in a frame. As

a result of this change in slot offset, the offset base generating section 2312 further increments the base offset value when the change of the offset of the slot from one input/output message packet 2230 to the next would extend beyond the number of slots in a frame. The slot count section 2313 generates the BUMP OFFSET BASE signal when this condition occurs.

The slot count section 2313 maintains a running index of the slot in the frame of the data item PE(x) MSG(y) for which the parallel send address/offset generator 2278(i) is currently generating DEST PE ADRS and DEST OFFSET signals. The slot count section 2313 uses the initial slot value and the slot incrementation value, as well as a correction value corresponding to the number of slots in a frame, to maintain the running index of the slot in the frame of the data item PE(x) MSG(y) for which the parallel send address/offset generator 2278(i) is currently generating DEST PE ADRS and DEST OFFSET signals. The initial slot value corresponds to the value of the index "i" of the parallel send address/offset generator 2278(i), modulo the number of slots in a frame. The slot incrementation value is, as noted above, the number "N" of input/output buffer modules 2201(i), modulo the number of slots in a frame. When the slot count section 2313 generates a slot count value that exceeds the number of slots in a frame, it asserts the BUMP OFFSET BASE signal and reduces the slot count value by the number of slots in a frame. The result is the offset of the slot in the next frame.

The destination processing element address generating section 2312 uses (i) the initial destination processing element address value, (ii) the destination processing element address incrementation value, (iii) the number of processing elements "NPE" participating in the input/output operation, (iv) the address of the first processing element 11(0) in the partition participating in the input/output operation, relative to the base of the processing element of the system, and (v) a BUMP DEST ADRS bump destination address signal from the offset delta generating section in generating DEST PE ADRS destination processing element address signals. The parallel send address/offset generator 2278(i) couples the DEST PE ADRS signals to the network interface 2277, which uses them in generating the message address portion 31 of the input/output message packet 2230.

It will be appreciated that, for the sequence of stripes in a frame, all of the data items PE(x) MSG(y) in slots in a stripe are to be transmitted in input/output message packets 2230 to one processing element 11(i) or scalar processor 12 participating in the input/output operation. The initial destination processing element address value for each parallel send address/offset generator 2278(i) thus relates to the stripe offset value for the stripe within the frame containing the first data item PE(x) MSG(y) to be transmitted by the input/output buffer node 2201(i). The stripe offset value, in turn, corresponds to the greatest integer of the quotient of the input/output buffer node's index "i" divided by the striping factor "C," modulo the number of stripes in a frame. The number of stripes in a frame corresponds to "NPE," the number of processing elements 11(i) and scalar processors 12 participating in the input/output operation.

The stripe offset value so generated is actually the offset, from the first processing element 11(0) or scalar processor 12 in the partition participating in the input/output operation, for the first input/output message packet 2230 to be generated by the input/output buffer node. Accordingly, the initial destination processing element address value is this stripe offset value plus the address of the first processing element 11(0) or scalar processor 12 participating in the input/output operation relative to the base of the processing element of the system 10.

The destination processing element address incrementation value is used by a parallel send address/offset generator 2278(i) when generating a destination processing element address for each subsequent input/output message packet 2230 generated by its input/output message buffer 2201(i). The destination processing element address incrementation value is related to the number of stripes within a frame that the input/output buffer node 2201(i) will increment between transmission of input/output message packets 2230. Thus, the destination processing element address incrementation value corresponds to the sum of the greatest integer of the number "N" of input/output buffer nodes 2201(i) participating in the input/output operation divided by the striping factor "C," modulo the number of stripes in a frame, that is, "NPE."

It will be appreciated that, if the number "N" of input/output buffer nodes 2201(i) participating in the input/output operation is not a multiple of the number of stripes in a frame, the offset of the slot containing the data item PE(x) MSG(y) being transmitted within a stripe will change for each subsequent input/output message packet. The change in the slot offset corresponds to the remainder of the quotient of (a) the number "N" of input/output buffer nodes 2201(i) participating in the input/output operation, divided by (b) the number of slots in a stripe, that is, the striping factor "C", which remainder, in turn, corresponds to the number "N" modulo the striping factor. As a result of this change in slot offset within a stripe, destination processing element address generating section 2310 further increments the destination processing element address when the change of the offset of the slot from one input/output message packet 2230 to the next would extend beyond the number of slots in a stripe. The offset delta generating section 2311 generates the BUMP DEST ADRS signal when this condition occurs.

The offset delta generating section 2311 also generates the OFFSET DELTA signal, which, as noted above, represents the component of the DEST OFFSET signal whose binary-encoded value identifies the position of the slot of the data item PE(x) MSG(y) being transmitted within a stripe, that is, within the series of data items within frame that are to be transmitted to the same processing element 11(i) or scalar processor 12. In addition, the offset delta generating section 2311 generates the BUMP DEST ADRS bump destination address signal which is directed to the destination processing element address generating section 2310.

The initial offset delta value for a parallel send address/offset generator 2278(i) corresponds to the offset of the slot containing the first data item PE(x) MSG(y) to be transmitted by the parallel send address/offset generator 2278(i) within the stripe. Thus, the initial offset delta value

corresponds to the remainder in the quotient of (a) the index "i" of input/output buffer node 2201(i), divided by (b) the number of slots in a frame, that is, the striping factor "C" times the number "NPE" of processing elements 11(i) participating in the input/output operation. Otherwise stated, the initial offset delta value corresponds to the input/output buffer node's index "i," modulo the striping factor "C".

The offset delta incrementation value is related to the number of slots within a stripe that the input/output buffer node 2201(i) will increment between transmission of input/output message packets 2230. As noted above, the number of stripes that the input/output buffer node 2201(i) will increment between transmission of input/output message packets 2230 is related to the change, if any, of the destination processing element address value as determined by the destination processing element address generating section 2310. Thus, the offset delta incrementation value is the remainder in the quotient of (a) the number "N" of input/output buffer nodes 2201(i) participating in the input/output operation, divided by (b) the number of slots in a stripe, that is, the striping factor "C." Otherwise stated, the offset delta incrementation value corresponds to the number "N" of input/output buffer nodes 2201(i) participating in the input/output operation, modulo the striping factor "C."

It will be appreciated that, if the incrementation of the offset delta value by the offset delta generating section 2311 from one input/output message packet 2230 to the next would result in an offset delta value greater than or equal to the striping factor "C," the offset delta value would actually relate to a slot in a stripe advanced beyond the stripe which is identified by the destination processing element address value as determined by the destination processing element address generating section 2310. This advanced stripe, in turn, includes slots whose data items PE(x) MSG(y) are to be transmitted to the next processing element 11(i) beyond that identified by the destination processing element address value. When that occurs, the offset delta generating section 2311 asserts the BUMP DEST ADRS bump destination address signal, to enable the destination processing element address generating section 2310 to further increment the destination processing element address. In addition, the offset delta generating section 2311 subtracts the striping factor from the incremented offset delta value, to point to the position of the slot, within the stripe associated with the destination processing element address generated by the destination processing element address generating section 2310 for the data item being transmitted, of the data item PE(x) MSG(y) being transmitted in the input/output message packet.

Similarly, at some point the destination processing element address generating section 2310 will increment the destination processing element address to be above the address of the highest-indexed processing element 11(i) or scalar processor 12 participating in the input/output operation. At that point, the destination processing element address generating section 2310 corrects the destination processing element address to a value which is the address of one of the processing elements or scalar processors participating in the transfer. In this operation, the destination

processing element address generating section 2310 reduces the incremented destination processing element address by an amount corresponding to NPE, the number of processing elements participating in the input/output operation. This will ensure that the destination processing element address points to a processing element or scalar processor participating in the input/output operation during the operation.

With this background, the structure and operation of parallel send address/offset generator 2278(i) will be described in connection with Figs. 4A and 4B. Initially, the initial destination processing element address, which is represented by block 2320, is coupled through multiplexer 2340 and latched in latch 2341. In addition, the destination processing element address increment value is stored in latch 2321 of the destination processing element address generating section 2310.

Similarly, the initial offset delta value and initial offset base value, which are represented by blocks 2322 and 2324, respectfully, are coupled through multiplexers 2342 and 2344, respectfully, as OFFSET DELTA and OFFSET BASE signals, respectfully. These signals are latched in latches 2343 and 2345, respectfully. They are also concurrently coupled to an adder 2314, which generates an OFF BASE + DEL offset base plus delta signal whose binary-encoded value represents the sum of binary-encoded value of the OFFSET DELTA and OFFSET BASE signals. The OFF BASE + DEL signal is latched in a latch 2315, which provides the DEST OFFSET destination offset signal.

Contemporaneously, the offset delta increment value and offset base increment value are stored in registers 2323 and 2325, respectively, of the offset delta generating section 2311 and offset base generating section 2312. The initial slot value, represented by block 2326, is coupled through multiplexer 2346 and stored in latch 2347, and the slot increment value is stored in register 2327 of the slot count section 2313.

In addition, various other values are stored in other registers. The destination processing element address generating section 2310, includes registers 2330 and 2331. As noted above, when incrementing to generate the destination processing element address values, at some point the incrementation may generate a value which represents a processing element address beyond the range of processing elements 11(i) or scalar processors 12 participating in the input/output operation. The value in register 2330 is used to assist in detecting such a condition.

As will be described below in connection with Fig. 4A, when incrementing the destination processing element address value, the destination processing element address generating section 2310 selects between the values in registers 2321 and 2331, depending on the relationship between the previously-determined destination processing element address value and the contents of register 2330. The value in register 2330 is used to determine when the destination processing element address value has been incremented to a point at which it would, when next incremented, identify a processing element 11(i) or scalar processor 12 beyond those participating in the input/output operation. Such a value corresponds to (a) the address of the last processing element 11(i) or scalar

processor 12 participating in the input/output operation, which is the address of the first processing element 11(0) or scalar processor 12 plus the number "NPE" of processing elements or scalar processors participating in the input/output operation, less (b) the amount by which it would be incremented, that is, the address increment value. If the destination processing element address generating section 2310 determines that the previously-determined destination processing element address value is less than the value stored in register 2330, the destination processing element address value, if incremented by the address increment value in register 2321, would remain in its permissible range. In that case, the destination processing element address generating section 2310 uses the value in register 2321 in the incrementation.

However, if the destination processing element address generating section 2310 determines that the previously-determined destination processing element address value is greater than or equal to the value in register 2330, if the destination processing element address value were incremented by the address increment value, it would be beyond its permissible range. In that case, as noted above, the incremented destination processing element address value is reduced by a value corresponding to the number "NPE" of processing elements and scalar processors participating in the input/output operation. The contents of register 2331 corresponds to the address increment value, reduced by the value "NPE." When this value is added to the previously-determined destination processing element address value, the result would be equivalent to reducing the incremented destination processing element address value by the value "NPE."

Similarly, the offset delta generating section 2311 includes two registers 2332 and 2333. As noted above, the offset delta value varies over a range relating to the striping factor, and the values in these registers are used to limit the offset delta value to that range. As will be described below in connection with Fig. 4A, when incrementing the offset delta value, the offset delta generating section 2311 selects between the values in registers 2323 and 2333, depending on the relationship between the previously-determined offset delta value and the contents of register 2332. The value in register 2332 is used to determine when the offset delta value has been incremented to a point at which it would, when next incremented, represent an offset delta value beyond its permissible range, that is, equal to or greater than the striping factor "C". Such a value corresponds to (a) the striping factor "C", less (b) the amount by which it would be incremented, that is, the offset delta increment value. If the offset delta generating section 2311 determines that the previously-determined offset delta value is less than the value stored in register 2332, the offset delta value, if incremented by the offset delta increment value in register 2323, would remain in its permissible range. In that case, the offset delta generating section 2311 uses the value in register 2323 in the incrementation.

However, if the offset delta generating section 2311 determines that the previously-determined offset delta value is greater than or equal to the value in register 2332, if the delta offset value were incremented by the delta increment value, it would be beyond its permissible range. In that case, as noted above, the incremented delta offset value is reduced by the striping factor "C" and

the BUMP DEST ADRS signal asserted to control the destination processing element address generating section 2310. The contents of register 2333 corresponds to the delta increment value, reduced by the striping factor "C." When this value is added to the previously-determined delta offset value, the result would be equivalent to reducing the incremented delta offset value by the striping factor "C."

The offset base generating section 2312 also has a register 2334 which stores a value corresponding to the offset base increment value plus the striping factor "C." The value in the register 2325 is used when the slot count section 2313 determines that the previously-incremented offset base value is to be incremented by the offset base increment value. On the other hand, the value in the register 2334 is used in the incrementation of the offset base value, which, as described above, is further incremented by an amount corresponding to the striping factor "C."

Finally, the slot count section 2313 includes two registers 2335 and 2336. Register 2335 stores a value which is used to determine when the slot index value has been incremented to a point at which it would, when next incremented, represent a slot index value beyond its permissible range, that is, equal to or greater than the number of slots in a frame, the striping factor "C" times the number "NPE" of processing elements 11(i) or scalar processors 12 participating in an input/output operation. The value in register 2335 is the striping factor "C" times the number "NPE," less the slot increment value. The value in register 2336 is the slot increment value less the number of slots in a frame.

As will be described below in connection with Fig. 4B, when incrementing the slot count value, the slot count section 2313 selects between the values in registers 2327 and 2336, depending on the relationship between the previously-determined slot count value and the contents of register 2335. The value in register 2335 is used to determine when the slot count value has been incremented to a point at which it would, when next incremented, identify a slot offset greater than the number of slots in a frame. Such a value corresponds to (a) the number of slots in a frame, which is the striping factor "C" times the number "NPE" of processing elements 11(i) and scalar processors 12 participating in the input/output operation, less (b) the slot increment value. If the slot count section 2313 determines that the previously-determined slot increment value is less than the value stored in register 2335, the slot increment value, if incremented by the slot increment value in register 2327, would remain in its permissible range. In that case, the slot count section 2313 uses the value in register 2327 in the incrementation.

However, if the slot count section 2313 determines that the previously-determined slot count value is greater than or equal to the value in register 2335, if the slot count value were incremented by the slot increment value, it would identify a slot beyond the end of the current frame. In that case, as noted above, the slot count section 2313 asserts the BUMP OFFSET BASE signal, to enable the offset base section 2312 to use the value in register 2334 in the incrementation of the offset base value. In addition, the slot count section 2313 generates an new slot count value whose value is

incremented by the slot increment value and reduced by a value corresponding to the number of slots in a frame. The contents of register 2331 corresponds to the slot increment value, reduced by the value corresponding to the number of slots in a frame. When this value is added to the previously-determined slot count value, the result would be equivalent to reducing the incremented slot count value by the value corresponding to the number of slots in a frame.

After the various registers have been loaded as described above, and enabled the initial values to be loaded into latches 2341, 2343, 2315, 2345 and 2347 for the initial input/output message packet 2230 to be generated by the input/output buffer node, the various sections 2310, 2311, 2312 and 2313 are enabled to concurrently perform a series of iterations to facilitate the generation of DEST PE ADRS signals and DEST OFFSET signals for use in connection with generation of input/output message packets 2230 for the subsequent data items PE(x) MSG(y) to be transmitted by the input/output buffer node.

With reference initially to Fig. 4A, in the offset delta generating section 2311, the LAT OFFSET DELTA latched offset delta signals from the latch 2343, which at this point have a binary-encoded value corresponding to the initial offset delta value, are coupled to one input terminal of an adder 2351. A second input terminal of adder 2351 receives a SEL OFFSET DELTA INC FACTOR selected offset delta increment factor signal from a multiplexer 2351. The adder 2350 generates INC OFF DEL incremented offset delta signals which are coupled as the OFFSET DELTA signal to the input terminal of latch 2343 and to one input terminal of adder 2314, which, in combination with the OFFSET BASE signal generated during the iteration by the offset base generating section 2312 as described below, will generate the DEST OFFSET destination offset signal. The INC OFF DEL signal from adder 2350 represents the incremented delta offset value for the iteration.

The SEL OFFSET DELTA INC FACTOR selected offset delta increment factor signal is provided by multiplexer 2351 under control of a comparator 2352. The comparator 2352, in turn, also receives the LAT OFFSET DELTA signal from latch 2343, as well as the signal from register 2332, and generates in response the BUMP DEST ADRS bump destination address signal. The comparator 2352 negates the BUMP DEST ADRS signal if it determines that the binary-encoded value of the LAT OFFSET DELTA signal is less than the value represented by the signal from the register 2332. When that occurs, the binary-encoded value of the LAT OFFSET DELTA signal, if incremented by adder 2350 by the offset delta increment value in register 2323, will remain within the permissible range of the offset delta value. Accordingly, the negated BUMP DEST ADRS signal enables the multiplexer to couple the signal from register 2323 as the SEL OFF DELTA INC FACTOR selected offset delta increment factor signal to adder 2350. The adder generates an INC OFF DEL incremented offset delta signal, which the multiplexer 2342 couples as the OFFSET DELTA signal to input terminals of latch 2343 and of adder 2314.

On the other hand, the comparator 2343 asserts the BUMP DEST ADRS signal if it determines that the binary-encoded value of the LAT OFFSET DELTA signal is greater than or

equal to the value represented by the signal from the register 2332. When that occurs, the binary-encoded value of the LAT OFFSET DELTA signal, if incremented by adder 2350 by the offset delta increment value in register 2323, will be beyond permissible range of the offset delta value. Accordingly, the asserted BUMP DEST ADRS signal enables the multiplexer to couple the signal from register 2333 as the SEL OFF DELTA INC FACTOR selected offset delta increment factor signal to adder 2350. Since, as noted above, the binary-encoded value of the signal from register 2333 corresponds to the delta increment value, reduced by the striping factor "C", when the adder generates an INC OFF DEL incremented offset delta signal, the binary-encoded value of the INC OFF DEL signal will be within the required range. The multiplexer 2342 couples the INC OFF DEL signal as the OFFSET DELTA signal to input terminals of latch 2343 and of adder 2314.

The destination processing element address generating section 2310 operates in a manner generally similar to the operation of the slot count section 2313. In destination processing element address generating section 2310, destination processing element address signals from the latch 2343, which at this point have a binary-encoded value corresponding to the initial destination processing element address value, are coupled to one input terminal of an adder 2352. A second input terminal of adder 2352 receives a SEL PE ADRS INCR FACTOR selected processing element address increment factor signal from a multiplexer 2353. Adder 2352 further has a carry input terminal "C" that is controlled by the BUMP DEST ADRS bump destination address signal. The adder 2352 generates an INC PE ADRS incremented processing element address signal which is coupled as to the input terminal of latch 2341. The INC PE ADRS signal from adder 2352 represents the incremented destination processing element address value for the iteration.

The SEL PE ADRS INCR FACTOR selected processing element address increment factor signal is provided by multiplexer 2353 under control of a comparator 2354 and multiplexer 2355. The comparator 2354, in turn, also receives the DEST PE ADRS destination processing element address signal from latch 2341, as well as the signal from register 2330. Comparator 2354 provides two output signals, including a RST IF GT reset if greater than signal and a RST IF GE reset if greater than or equal to signal. The comparator 2354 asserts the RST IF GT signal if the binary-encoded value of the DEST PE ADRS signal is greater than the binary-encoded value of the signal from register 2330. On the other hand, the comparator asserts the RST IF GE signal if the binary-encoded value of the DEST PE ADRS signal is greater than or equal to the binary-encoded value of the signal from register 2330. Thus, comparator 2354 asserts the RST IF GE signal, but not the RST IF GT signal, if the binary-encoded value of the DEST PE ADRS signal corresponds to the value stored in register 2330.

The multiplexer 2355, under control of the BUMP DEST ADRS bump destination address signal, selectively couples one of the RST IF GE or RST IF GT signals as a RST PE ADRS reset processing element address signal to control multiplexer 2353. If the offset delta generating section 2311 is asserting the BUMP DEST ADRS signal, the multiplexer 2355 couples the RST IF GT reset

if greater than signal to the multiplexer 2353 as the RST PE ADRS reset processing element address signal. On the other hand, if the offset delta generating section 2311 is negating the BUMP DEST ADRS signal, the multiplexer 2355 couples the RST IF GE reset if greater than or equal to signal as the RST PE ADRS signal.

The multiplexer 2355 ensures that, when the destination processing element address generating section 2310 uses the BUMP DEST ADRS bump destination address signal, which is coupled to the carry in input terminal C_i of the adder 2352, to further increment the destination processing element address value, it does not increment the value beyond the permissible range of destination processing element address values. If the BUMP DEST ADRS signal is negated, so that the destination processing element address value will not be further incremented thereby, multiplexer 2355 couples the RST IF GT reset if greater than signal as an RST PE ADRS reset processing element address signal. Under this condition, if the comparator 2355 determines that the binary-encoded value of the DEST PE ADRS destination processing element address signal is less than or equal to the binary-encoded value of the signal from register 2330, the RST IF GT signal will be negated. The negated BUMP DEST ADRS signal will enable multiplexer 2355 to couple the negated RST IF GT signal to the multiplexer 2353, which, in turn, enables the multiplexer 2353 to couple an SEL PE ADRS INC FACTOR selected processing element address increment factor signal representing the address increment value to the second input terminal of adder 2352. Adder 2352 generates an INC PE ADRS incremented processing element address signal representing the sum of the binary-encoded values of the DEST PE ADRS signal, the SEL PE ADRS INC FACTOR signal, which the multiplexer 2340 couples the INC PE ADRS signal to the input terminal of latch 2341.

If, however, while the BUMP DEST ADRS signal is negated the comparator 2355 determines that binary-encoded value of the DEST PE ADRS signal is greater than the binary-encoded value of the signal from register 2330, the RST IF GT signal will be asserted. In that case, the RST PE ADRS signal will also be asserted, enabling the multiplexer 2353 to couple an SEL PE ADRS INC FACTOR selected processing element address increment factor signal corresponding to the address increment value reduced by the value "NPE," to the second input terminal of adder 2352. Adder 2352 generates an INC PE ADRS incremented processing element address signal representing the sum of the binary-encoded values of the DEST PE ADRS signal and the SEL PE ADRS INC FACTOR signal. The multiplexer 2340 couples the INC PE ADRS signal to the input terminal of latch 2341.

If, on the other hand, the BUMP DEST ADRS signal is asserted, the adder 2352 will generate INC PE ADRS incremented processing element address signals whose binary-encoded value corresponds to the sum of the binary-encoded values of the DEST PE ADRS destination processing element address signals and the SEL PE ADRS INC FACTOR selected processing element address increment factor, as further incremented since the BUMP DEST ADRS signal is

asserted. In that case, to ensure that the adder 2352 does not increment the DEST PE ADRS signal to provide a destination processing element address beyond that for the processing elements 11(i) and scalar processors 12 participating in the input/output operation, the BUMP DEST ADRS signal enables the multiplexer 2355 to couple the RST IF GE reset if greater than or equal to signal as the RST PE ADRS signal.

Accordingly, if the comparator 2355 determines that the binary-encoded value of the DEST PE ADRS destination processing element address signal is less than the binary-encoded value of the signal from register 2330, the RST IF GE signal will be negated. The asserted BUMP DEST ADRS signal will enable multiplexer 2355 to couple the negated RST IF GE signal to the multiplexer 2353, which, in turn, enables the multiplexer 2353 to couple an SEL PE ADRS INC FACTOR selected processing element address increment factor signal representing the address increment value to the second input terminal of adder 2352. Adder 2352 generates an INC PE ADRS incremented processing element address signal representing the sum of the binary-encoded values of the DEST PE ADRS signal, the SEL PE ADRS INC FACTOR signal, along with the asserted BUMP DEST ADRS signal as applied to its carry in terminal C_i, which the multiplexer 2340 couples the INC PE ADRS signal to the input terminal of latch 2341.

If, however, while the BUMP DEST ADRS signal is asserted the comparator 2355 determines that binary-encoded value of the DEST PE ADRS signal is greater than or equal to the binary-encoded value of the signal from register 2330, the RST IF GE signal will be asserted. In that case, the RST PE ADRS signal will also be asserted, enabling the multiplexer 2353 to couple an SEL PE ADRS INC FACTOR selected processing element address increment factor signal corresponding to the address increment value reduced by the value "NPE," to the second input terminal of adder 2352. Adder 2352 generates an INC PE ADRS incremented processing element address signal representing the sum of the binary-encoded values of the DEST PE ADRS signal, the SEL PE ADRS INC FACTOR signal, along with the BUMP DEST ADRS signal at its carry-in input terminal C_i. The multiplexer 2340 couples the INC PE ADRS signal to the input terminal of latch 2341.

With reference to Fig. 4B, in the slot count section 2313, the LAT SLOT INDEX latched slot index signal from the latch 2347, which at this point have a binary-encoded value corresponding to the initial slot index value, are coupled to one input terminal of an adder 2360. A second input terminal of adder 2360 receives a SEL SLOT INDEX INC FACTOR selected slot index increment factor signal from a multiplexer 2361. The adder 2360 generates an INC SLOT INDEX incremented slot index signal which multiplexer 2346 couple as a SLOT INDEX signal to the input terminal of latch 2343. The SEL SLOT INDEX INC FACTOR selected slot index increment factor signal is provided by multiplexer 2361 under control of a comparator 2362.

The comparator 2362, in turn, also receives the LAT SLOT INDEX signal from latch 2347, as well as the signal from register 2335, and generates in response the BUMP OFFSET BASE bump offset base signal. The comparator 2362 negates the BUMP OFFSET BASE signal if it determines

that the binary-encoded value of the LAT slot index signal is less than the value represented by the signal from the register 2335. When that occurs, the binary-encoded value of the LAT SLOT INDEX signal, if incremented by adder 2360 by the slot increment value in register 2327, will remain within the permissible range of the slot index value. Accordingly, the negated BUMP OFFSET BASE signal enables the multiplexer 2361 to couple the signal from register 2327 as the SEL SLOT INDEX INC FACTOR selected slot index increment factor signal to adder 2360. The adder generates an INC SLOT INDEX incremented slot index signal, which the multiplexer 2346 couples as the SLOT INDEX signal to input terminals of latch 2347.

On the other hand, the comparator 2362 asserts the BUMP SLOT INDEX signal if it determines that the binary-encoded value of the LAT OFFSET DELTA signal is greater than or equal to the value represented by the signal from the register 2335. When that occurs, the binary-encoded value of the LAT SLOT INDEX signal, if incremented by adder 2360 by the increment slot index value in register 2327, will be beyond permissible range of the slot index value. Accordingly, the asserted BUMP OFFSET BASE signal enables the multiplexer 2361 to couple the signal from register 2336 as the SEL SLOT INDEX INC FACTOR selected slot index increment factor signal to adder 2360. Since, as noted above, the binary-encoded value of the signal from register 2336 corresponds to the slot increment value, reduced by the number of slots in a frame, when the adder 2360 generates an INC SLOT INDEX incremented slot index signal, the binary-encoded value of the INC slot signal will be within the required range. The multiplexer 2346 couples the INC SLOT INDEX signal as the SLOT INDEX signal to the input terminal of latch 2347.

In the offset base generating section 2312, the LAT OFFSET BASE latched offset base signal from the latch 2345, which at this point has a binary-encoded value corresponding to the initial offset base value, is coupled to one input terminal of an adder 2363. A second input terminal of adder 2363 receives a SEL OFF BASE INC FACTOR selected offset base increment factor signal from a multiplexer 2364. The adder 2363 generates an INC OFF BASE incremented offset base signal which multiplexer 2344 couples as the offset base signal to the input terminal of latch 2345 and to one input terminal of adder 2314. As described above, adder 2314 generates an OFFSET BASE + DEL offset base plus delta signal, whose binary-encoded value corresponds to the sum of the binary-encoded values of the OFFSET BASE and OFFSET DELTA signals, and which is coupled to the input terminal of latch 2315.

The SEL OFF BASE INC FACTOR selected offset base increment factor signal is provided by multiplexer 2364 under control of the BUMP OFFSET BASE signal from comparator 2362. As described above, the comparator 2362 negates the BUMP OFFSET BASE signal if it determines that the binary-encoded value of the LAT slot index signal is less than the value represented by the signal from the register 2335. When that occurs, the binary-encoded value of the LAT SLOT INDEX signal, if incremented by adder 2360 by the slot increment value in register 2327, will remain within the permissible range of the slot index value. In that case, the negated BUMP OFFSET BASE signal

enables the multiplexer 2364 to couple the signal from register 2325, representing the offset base increment value, as the SEL OFF BASE INC FACTOR selected offset base increment factor signal to adder 2363. The adder 2363 generates an INC OFF BASE incremented offset base signal, which the multiplexer 2346 couples as the OFFSET BASE signal to input terminals of latch 2347 and adder 2314.

On the other hand, the comparator 2362 asserts the BUMP SLOT INDEX signal if it determines that the binary-encoded value of the LAT OFFSET DELTA signal is greater than or equal to the value represented by the signal from the register 2335. When that occurs, the binary-encoded value of the LAT SLOT INDEX signal, if incremented by adder 2363 by the increment slot index value in register 2327, will be beyond permissible range of the slot index value. Accordingly, the asserted BUMP OFFSET BASE signal enables the multiplexer 2364 to couple the signal from register 2334, representing the offset base increment value plus the striping factor "C" as the SEL OFF BASE INC FACTOR selected offset base increment factor signal, to adder 2363. In that case, adder 2363 generates an INC OFF BASE incremented offset base signal whose a binary-encoded value corresponds to the binary-encoded value of the LAT OFFSET BASE signal, incremented by both the offset base increment value and the striping factor "C."

As noted above, the various sections 2310, 2312, 2312 and 2313 of the parallel send address/offset generator 2278(i) iteratively perform these operations to generate the DEST PE ADRS destination processing element address signals and DEST OFFSET destination offset signals to be used in connection with generation of the input/output message packets 2230. During each iteration, the input/output message packet 2230 transmitted by the input/output buffer 2201(i) includes one data item PE(x) MSG(y) from of its buffer memory 2223(i). After the input/output buffer 2201(i) has transmitted all of the data items PE(x) MSG(y) it may terminate the input/output operation.

It will be appreciated that numerous modifications may be made the parallel send address/offset generator 2278(i) described above. For example, instead of providing separate adders and comparators for the various sections 2310, 2311, 2312 and 2313, the parallel send address/offset generator may have a single adder and comparator, which may be shared among the various sections. In such an embodiment, the adder and comparator would be used in separate phases, during each phase to generate signals representing the destination processing element address value, offset delta value, offset base value and slot index value. In that case, the adder and comparator would be used to generate the offset delta value before the destination processing element address value, since they will require the BUMP DEST ADRS signal to generate the destination processing element address value. In addition, the adder and comparator would be used to generate the slot index value before the destination base value, since they will require the BUMP OFFSET BASE signal to generate the offset base value. Such an embodiment may be useful in reducing the physical size of the circuit comprising the parallel send address/offset generator 2278(i), although it will be appreciated that it

may require more time to generate the destination processing element address value and destination offset value since they are determined in a four phase sequence.

In addition, it will be appreciated that the destination processing element address value and destination offset value may be determined using a suitably-programmed microprocessor.

The foregoing description has been limited to a specific embodiment of this invention. It will be apparent, however, that variations and modifications may be made to the invention, with the attainment of some or all of the advantages of the invention. Therefore, it is the object of the appended claims to cover all such variations and modifications as come within the true spirit and scope of the invention.

What is claimed as new and desired to be secured by Letters Patent is:

CLAIMS

1. A computer comprising a plurality of processing elements, and an input/output processor interconnected by a routing network,

A. said routing network transferring messages between said processing elements and said input/output processor;

B. said processing elements performing processing operations in connection with data received from said input/output processor in messages transferred over said routing network and transferring processed data to said input/output processor in messages over said routing network, said processing elements being connected as a first selected series of leaf nodes;

C. said input/output processor including a plurality of input/output buffers connected as a second selected series of leaf nodes of said routing network for generating messages for transfer over said routing network to a series of processing elements forming at least a selected subset of the processing elements during an input/output operation.

2. A computer as defined in claim 1 in which said input/output processor further receives messages over said routing network from a series of processing elements forming at least a selected subset of the processing elements during an input/output operation.

3. A computer as defined in claim 1 further comprising at least one control processor and a control network, said control processor generating processing control messages for transfer to said processing elements over said control network to control said processing elements.

4. A computer as defined in claim 3 comprising a plurality of control processors each generating processing control messages for transfer to at least selected ones of said processing elements over said control network to control said processing elements, said control network being partitionable to define a plurality of partitions each facilitating the transfer of processing control messages between at least one control processor and selected ones of said processing elements.

5. A computer as defined in claim 3 in which said control processor further generates input/output control messages and said input/output processor further includes a common control for receiving said input/output control messages and controlling said input/output buffers to perform input/output operations in response thereto.

6. A computer as defined in claim 5 in which each processing element in a selected subset is identified by an address and includes a processing element receive buffer for buffering data from messages received from said routing network during an input/output operation, each processing element buffering data received in a message at an offset in said processing element buffer identified by a destination offset value in the message, each input/output buffer including:

A. a transmit buffer for buffering a plurality of data items, each to be transmitted in a message to a processing element; and

B. a destination processing element address and offset generator for iteratively generating a destination processing element address value and a destination offset value.

7. A computer as defined in claim 6 in which the transmit buffer of each input/output buffer includes a plurality of storage locations at a series of source offsets, each storage location storing a data item used a message, the data items defining a succession of frames of storage locations in the transmit buffers of said input/output buffers, each frame being organized first in order of successive input/output buffers in storage locations having the same source offset and second in order of storage locations in each input/output buffer having successive source offsets so as to include data items to be received by the series of processing elements participating in the input/output operation for storage in their respective processing element receive buffers at the same destination offset value, the destination processing element address and offset generator iteratively generating destination processing element address values and destination offset values in response to the number of input/output buffers and the number of processing elements participating in the input/output operation.

8. A computer as defined in claim 7 in which the destination processing element address and offset generator further generates during an initial iteration an initial destination processing element address value and an initial destination offset value both related to the number of input/output buffers, the number of processing elements participating in the input/output operation, and the position of the input/output buffer among the input/output buffers participating in the input/output operation, the destination processing element address and offset generator during subsequent iterations generating a destination processing element address value and destination offset value in response to the initial destination processing element address value and an initial destination offset value

9. A computer as defined in claim 7 in which said destination processing element address and offset generator further generates said destination processing element address value in response to a base processing element address value identifying a predetermined one of the processing elements in the series of processing elements participating in the input/output operation.

10. A computer as defined in claim 7 in which said destination processing element address and offset generator comprises:

A. a destination processing element address value generator for, during successive iterations, generating destination processing element address values in response to an initial destination processing element address value, the number of input/output buffers and the number of processing elements participating in the input/output operation said destination processing element address value generating during an iteration identifying the data item to be used in a message during the iteration within the sequence of data items comprising its frame; and

B. a destination offset value generator for, during successive iterations, generating destination offset values in response to an initial destination offset value, the number of input/output buffers and the number of processing elements participating in the input/output operation, the destination offset value generated during an iteration identifying the frame containing the data item to be used in a message during the iteration within the sequence of frames to be transferred.

11. A computer as defined in claim 10 in which said destination processing element address value generator includes:

A. a destination processing element address value store for storing a destination processing element address value;

B. an address incrementation value store for storing an address incrementation value; and

C. a destination address value incrementation circuit for generating, during each iteration, an incremented destination processing element address value in response to the destination processing element address value stored in said destination processing element address value store and the address incrementation value, the incremented destination processing element address value being stored in the destination processing element address value store as the destination processing element address value for use during the next iteration.

12. A computer as defined in claim 11 in which the address incrementation value stored in the address incrementation value store is related to the number of processing elements and the number of input/output buffers participating in the input/output operation.

13. A computer as defined in claim 11 wherein said destination processing element address value generator further includes an destination address initialization circuit for enabling the destination processing element address value store to store an initial destination offset value both related to the number of input/output buffers, the number of processing elements participating in the input/output operation, and the position of the input/output buffer among the input/output buffers participating in the input/output operation.

14. A computer as defined in claim 11 wherein said destination address value incrementation circuit further includes a destination processing element address value range limitation circuit for limiting incremented destination processing element address value to an address value range corresponding to the address values of the processing elements participating in the input/output operation.

15. A computer as defined in claim 14 in which:

A. said destination processing element address value incrementation circuit further includes:

i. destination processing element address value range limitation store for storing a limitation value relating to an upper end of the address value range; and

ii. an address reset store for storing an address reset value;

B. said destination processing element address value range limitation circuit includes:

i. a selector circuit for selectively coupling either the address incrementation value from said address incrementation value store or the address reset value from said address reset store to said destination address value incrementation circuit in response to a selection control signal; and

ii. a comparator for generating said selection control signal in response to the destination processing element address value from said destination processing element address value store and the limitation value from said destination processing element address value range limitation store, the address reset value and the limitation value being selected to ensure that the incremented

destination processing element address value generated by said destination address value incrementation circuit is within said address value range.

16. A computer as defined in claim 10 in which said destination offset value generator includes:

A. a destination offset value store for storing a destination offset value;

B. an offset incrementation value store for storing an offset incrementation value; and

C. a destination offset value incrementation circuit for generating, during each iteration, an incremented destination offset value in response to the destination offset value stored in said destination offset value store and the offset incrementation value, the incremented destination offset value being stored in the destination processing element offset value store as the destination offset value for use during the next iteration.

17. A computer as defined in claim 16 in which the offset incrementation value stored in the offset incrementation value store is related to the number of processing elements and the number of input/output buffers participating in the input/output operation.

18. A computer as defined in claim 16 wherein said destination offset value generator further includes an destination offset initialization circuit for enabling the destination offset value store to store an initial destination offset value related to the number of processing elements participating in the input/output operation and the position of the input/output buffer among the input/output buffers participating in the input/output operation.

19. A computer as defined in claim 7 in which each frame is further defined as including a series of stripes, the series including data items each to be received by the series of processing elements participating in the input/output operation, each stripe including a predetermined number of data items to be received by the series of processing elements participating in the input/output operation for storage in their respective processing element receive buffers at successive destination offset values, said destination processing element address value generator further generating said destination processing element address values and destination offset values in response to the number of data items in each stripe.

20. A computer as defined in claim 19 in which the destination processing element address and offset generator further generates during an initial iteration an initial destination processing element address value and an initial destination offset value both related to the number of input/output buffers, the number of processing elements participating in the input/output operation, the position of the input/output buffer among the input/output buffers participating in the input/output operation, and the number of data items in each stripe, the destination processing element address and offset generator during subsequent iterations generating a destination processing element address value and destination offset value in response to the initial destination processing element address value and an initial destination offset value

21. A computer as defined in claim 19 in which said destination processing element address and

offset generator further generates said destination processing element address value in response to a base processing element address value identifying a predetermined one of the processing elements in the series of processing elements participating in the input/output operation.

22. A computer as defined in claim 19 in which said destination processing element address and offset generator comprises:

A. a destination offset value generator for, during successive iterations, generating destination offset values in response to an initial destination offset value, the number of input/output buffers, the number of processing elements participating in the input/output operation, and the number of data items in a stripe, said destination offset value for each iteration identifying the frame and the position of the data item to be used in a message during the iteration within the sequence of data items comprising its stripe, the destination offset value generator further generating a destination address control signal having selected conditions; and

B. a destination processing element address value generator for, during successive iterations, generating destination processing element address values in response to an initial destination processing element address value, the number of input/output buffers, the number of processing elements participating in the input/output operation, and the condition of the destination address control signal, the destination processing element address value for each iteration identifying the stripe which contain the data item to be used in a message during the iteration within the sequence of stripes comprising a frame.

23. A computer as defined in claim 22 in which said destination processing element address value generator includes:

A. a destination processing element address value store for storing a destination processing element address value;

B. an address incrementation value store for storing an address incrementation value; and

C. a destination address value incrementation circuit for generating, during each iteration, an incremented destination processing element address value in response to the destination processing element address value stored in said destination processing element address value store, the address incrementation value, and the condition of the destination address control signal, the incremented destination processing element address value being stored in the destination processing element address value store as the destination processing element address value.

24. A computer as defined in claim 23 in which the address incrementation value stored in the address incrementation value store is related to the number of processing elements and the number of input/output buffers participating in the input/output operation.

25. A computer as defined in claim 23 wherein said destination processing element address value generator further includes an destination initialization circuit for enabling the destination processing element address value store to store an initial destination offset value both related to the number of input/output buffers, the number of processing elements participating in the input/output operation,

the position of the input/output buffer among the input/output buffers participating in the input/output operation, and the number of data items in a stripe.

26. A computer as defined in claim 23 wherein said destination address value incrementation circuit further includes a destination processing element address value range limitation circuit for limiting incremented destination processing element address value to an address value range corresponding to the address values of the processing elements participating in the input/output operation.

27. A computer as defined in claim 26 in which:

A. said destination processing element address value incrementation circuit further includes:

i. destination processing element address value range limitation store for storing a limitation value relating to an upper end of the address value range; and

ii. an address reset store for storing an address reset value;

B. said destination processing element address value range limitation circuit includes:

i. a selector circuit for selectively coupling either the address incrementation value from said address incrementation value store or the address reset value from said address reset store to said destination address value incrementation circuit in response to a selection control signal; and

ii. a comparator for generating said selection control signal in response to the destination processing element address value from said destination processing element address value store and the limitation value from said destination processing element address value range limitation store, the address reset value and the limitation value being selected to ensure that the incremented destination processing element address value generated by said destination address value incrementation circuit is within said address value range.

28. A computer as defined in claim 22 in which said destination offset value generator includes:

A. a destination offset base value generator for generating a destination base offset value during each iteration, said destination base offset value identifying the frame containing the data item to be used in a message during the iteration;

B. a destination offset delta generator for generating a destination delta offset value during each iteration, the destination delta offset value identifying the position of the data item to be used in a message during the iteration within the sequence of data items comprising its stripe; and

C. a destination offset combination value generator for generating said destination offset value in response to said destination base offset value and said destination delta offset value.

29. A computer as defined in claim 28 in which said destination offset base value generator comprises:

A. a destination base offset value store for storing a destination base offset value to be used by the destination offset combination value generator;

B. a destination base offset value incrementation circuit for generating, during each iteration, an incremented destination base offset value in response to the destination base offset value stored in said destination base offset value store and a base offset incrementation value, the

incremented destination base offset value being stored in the destination base offset value store as the destination offset base value for use during the next iteration;

C. a base offset incrementation value circuit for providing a base offset incrementation value, said base offset incrementation value circuit comprising:

i. a base offset base incrementation value store for storing a base incrementation value;

ii. base offset enhanced incrementation value store for storing an enhanced incrementation value reflecting the base incrementation value and the number of data items in a stripe; and

iii. base offset incrementation value selector for selectively coupling one of the base incrementation value or the enhanced incrementation value as the base offset incrementation value in response to a slot signal; and

D. a slot count circuit for maintaining a running count of the data item to be transmitted in a message during an iteration in a sequence of data items within a frame and generating the slot count signal in response to the running count and the number of data items within a frame.

30. A computer as defined in claim 29 wherein said destination base offset value generator further includes a destination base offset initialization circuit for enabling the destination base offset value store to store an initial destination base offset value related to the number of processing elements participating in the input/output operation, the position of the input/output buffer among the input/output buffers participating in the input/output operation, and the number of data items in a stripe.

31. A computer as defined in claim 29 in which said slot count circuit comprises:

A. a slot count store for storing a slot count value;

B. a slot count incrementation value store for storing an slot count incrementation value; and

C. a slot count incrementation circuit for generating, during each iteration, an incremented slot count value in response to the slot count value stored in said slot count store and the slot count incrementation value, the incremented slot count value being stored in the slot count store as the slot count value to be used in the next iteration.

32. A computer as defined in claim 31 in which the slot incrementation value stored in the slot count incrementation value store is related to the number of processing elements and the number of input/output buffers participating in the input/output operation, and the number of data items in a stripe.

33. A computer as defined in claim 31 wherein said slot count circuit further includes a slot count initialization circuit for enabling the slot count store to store an initial slot count value related to the number of processing elements participating in the input/output operation, the position of the input/output buffer among the input/output buffers participating in the input/output operation, and the number of data items in a stripe.

34. A computer as defined in claim 31 wherein said slot count circuit further includes a slot count

value range limitation circuit for limiting the incremented slot count value to a slot count value range corresponding to the number of data items in a frame.

35. A computer as defined in claim 28 wherein said destination offset delta generator comprises:

A. a destination delta offset value store for storing a destination delta offset value to be used by the destination offset combination value generator;

B. a destination delta offset value incrementation circuit for generating, during each iteration, an incremented destination delta offset value in response to the destination delta offset value stored in said destination delta offset value store and a delta offset incrementation value, the incremented destination delta offset value being stored in the destination delta offset value store as the destination delta offset value for use during the next iteration;

C. a delta offset incrementation value circuit for providing a base offset incrementation value, said base offset incrementation value circuit comprising:

i. a delta offset base incrementation value store for storing a delta incrementation value;

ii. delta offset reduced incrementation value store for storing a reduced delta incrementation value reflecting the delta incrementation value and the number of data items in a stripe; and

iii. a delta offset incrementation value selector for selectively coupling one of the delta incrementation value or the reduced delta incrementation value as the base offset incrementation value in response to the destination delta offset value and the number if data items in a stripe.

36. A computer as defined in claim 35 wherein said destination delta offset value generator further includes a destination delta offset initialization circuit for enabling the destination delta offset value store to store an initial destination delta offset value related to the position of the input/output buffer among the input/output buffers participating in the input/output operation and the number of data items in a stripe.

37. A computer as defined in claim 35 wherein said destination delta offset value generator further includes a destination delta offset value range limitation circuit for generating said destination address control signal and for limiting the incremented delta offset value to a delta offset value range corresponding to number of data items in a stripe.

38. A computer comprising:

A. a plurality of processing elements for performing processing operations in accordance with processing control messages to generate processed data in connection with data received in data messages and for generating data messages containing said processed data;

B. a plurality of control processors for generating said processing control messages for controlling processing by said processing elements and for generating input/output control messages;

C. an input/output processor responsive to input/output control messages from said control processors for initiating an input/output operation to transfer data in data messages with at least a selected subset of said processing elements;

D. a routing network for transferring data messages between said processing elements and

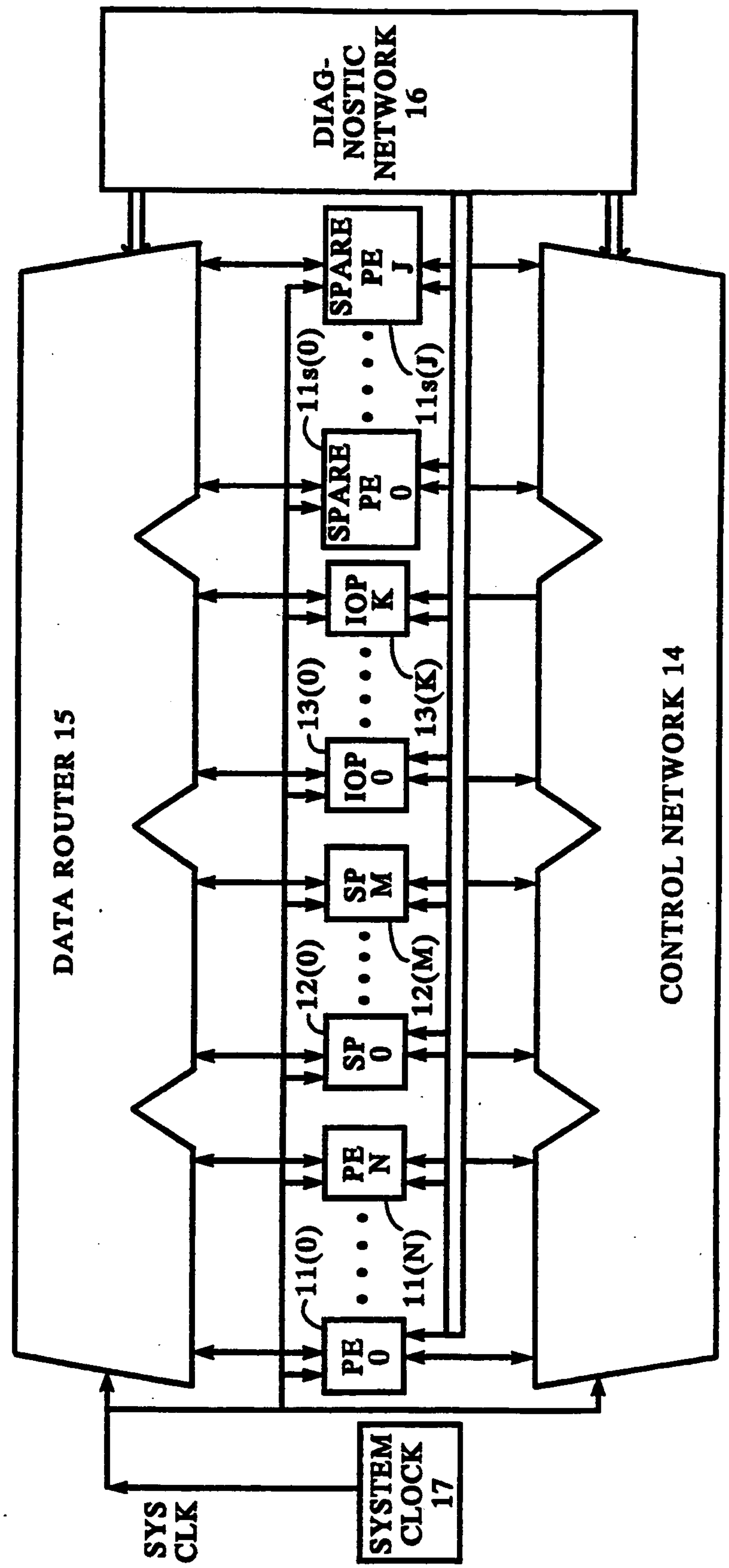
said input/output processor and input/output control messages between said control processors and said input/output processor; and

E. a control network for transferring said processing control messages between said control processors and said processing elements, said control network being partitionable into a plurality of partitions each facilitating the transfer of processing control messages between at least one control processor and selected ones of said processing elements.

39. An input/output processor including a plurality of input/output buffers connected to a series of leaf nodes of said routing network for generating messages for transfer over said routing network to a plurality of data receivers each connected to one of a second series of nodes of said routing network and identified by an address during an input/output operation, each input/output buffer including:

A. a transmit buffer for buffering a plurality of data items each to be transmitted in a message to a data receiver in a message; and

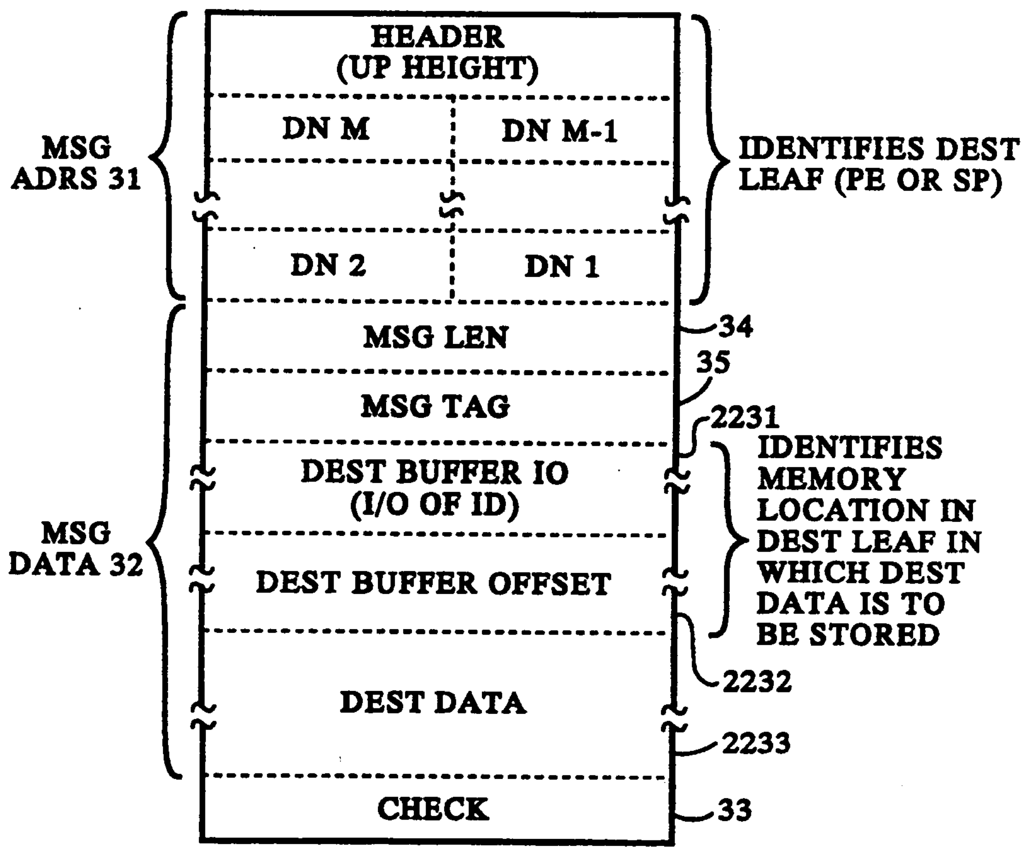
B. a destination data receiver address and offset generator for iteratively generating a destination data receiver address value and a destination offset value in response to the number of input/output buffers and the number of data receivers participating in the input/output operation.



SYSTEM 10

FIG. 1

INPUT / OUTPUT
MESSAGE PACKET 2230



I/O PROCESSOR
I/O MESSAGE FORMAT

FIG. 2

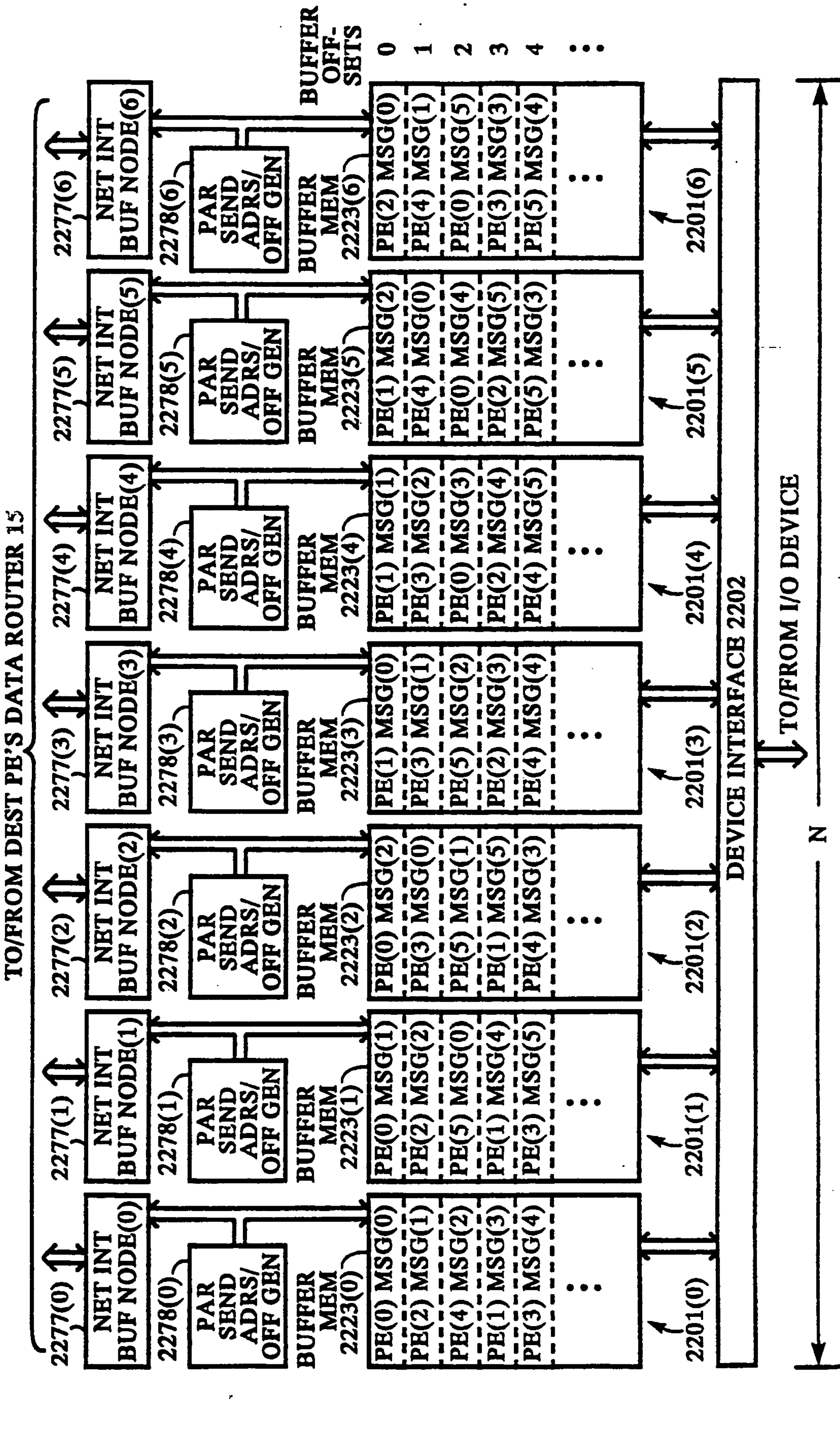

TO/FROM DEST PE'S DATA ROUTER 15
2277(0)
NET INT BUF NODE(0)
2278(0)
PAR SEND ADRS/ OFF GEN
BUFFER MEM 2223(0)
PE(0) MSG(0)
PE(2) MSG(1)
PE(4) MSG(2)
PE(1) MSG(3)
PE(3) MSG(4)
2201(0)
2277(1)
NET INT BUF NODE(1)
2278(1)
PAR SEND ADRS/ OFF GEN
BUFFER MEM 2223(1)
PE(0) MSG(1)
PE(2) MSG(2)
PE(5) MSG(0)
PE(1) MSG(4)
PE(3) MSG(5)
2201(1)
2277(2)
NET INT BUF NODE(2)
2278(2)
PAR SEND ADRS/ OFF GEN
BUFFER MEM 2223(2)
PE(0) MSG(2)
PE(3) MSG(0)
PE(5) MSG(1)
PE(1) MSG(5)
PE(4) MSG(3)
2201(2)
2277(3)
NET INT BUF NODE(3)
2278(3)
PAR SEND ADRS/ OFF GEN
BUFFER MEM 2223(3)
PE(1) MSG(0)
PE(3) MSG(1)
PE(5) MSG(2)
PE(2) MSG(3)
PE(4) MSG(4)
2201(3)
2277(4)
NET INT BUF NODE(4)
2278(4)
PAR SEND ADRS/ OFF GEN
BUFFER MEM 2223(4)
PE(1) MSG(1)
PE(3) MSG(2)
PE(0) MSG(3)
PE(2) MSG(4)
PE(4) MSG(5)
2201(4)
2277(5)
NET INT BUF NODE(5)
2278(5)
PAR SEND ADRS/ OFF GEN
BUFFER MEM 2223(5)
PE(1) MSG(2)
PE(4) MSG(0)
PE(0) MSG(4)
PE(2) MSG(5)
PE(5) MSG(3)
2201(5)
2277(6)
NET INT BUF NODE(6)
2278(6)
PAR SEND ADRS/ OFF GEN
BUFFER MEM 2223(6)
PE(2) MSG(0)
PE(4) MSG(1)
PE(0) MSG(5)
PE(3) MSG(3)
PE(5) MSG(4)
2201(6)
BUFFER OFF-SETS
0
1
2
3
4
DEVICE INTERFACE 2202
TO/FROM I/O DEVICE
N

FIG. 3A

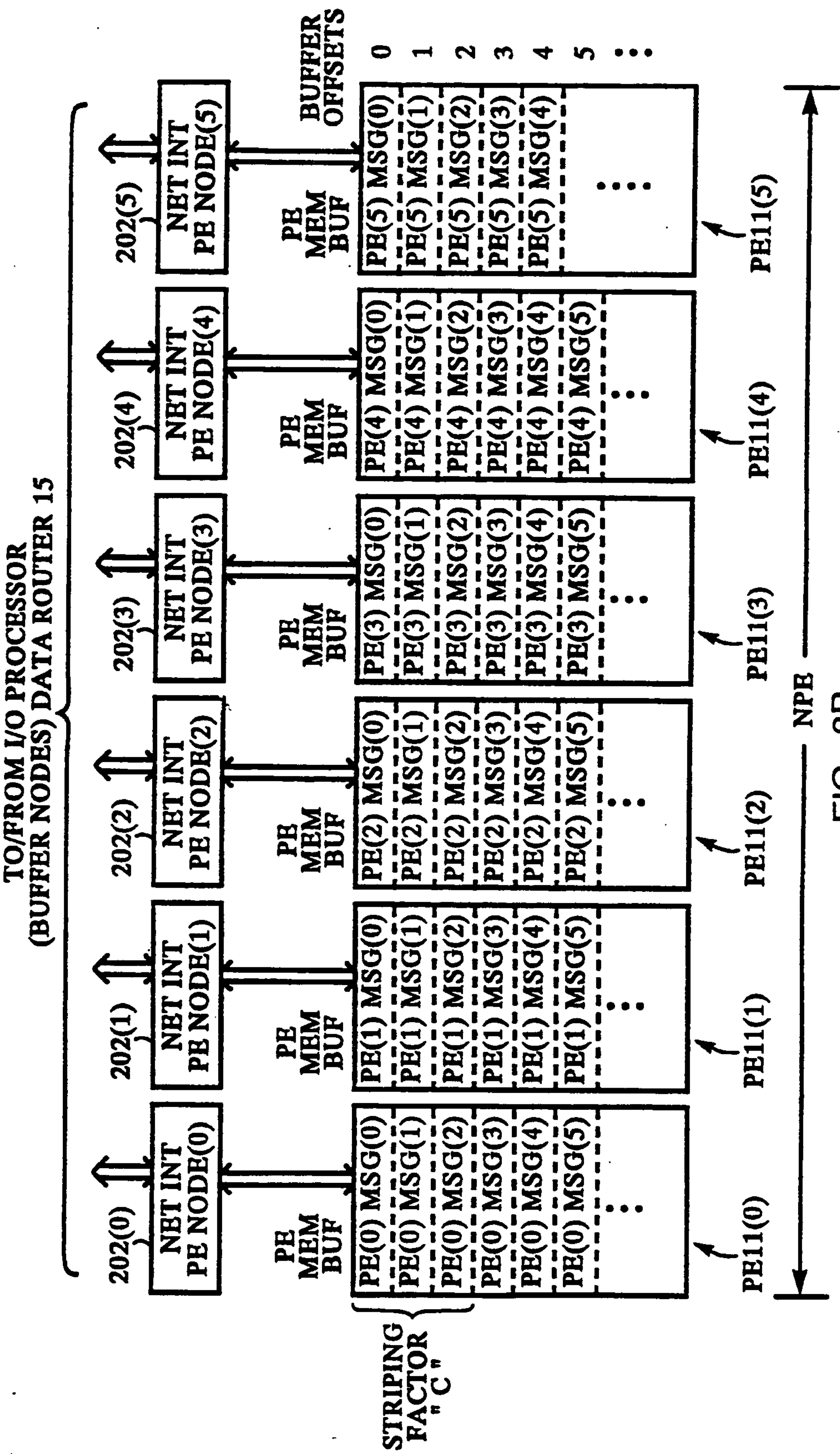

TO/FROM I/O PROCESSOR
(BUFFER NODES) DATA ROUTER 15
202(0)
NET INT
PE NODE(0)
202(1)
NET INT
PE NODE(1)
202(2)
NET INT
PE NODE(2)
202(3)
NET INT
PE NODE(3)
202(4)
NET INT
PE NODE(4)
202(5)
NET INT
PE NODE(5)
PE
MEM
BUF
BUFFER
OFFSETS
0
1
2
3
4
5
PE(0) MSG(0)
PE(0) MSG(1)
PE(0) MSG(2)
PE(0) MSG(3)
PE(0) MSG(4)
PE(0) MSG(5)
PE(1) MSG(0)
PE(1) MSG(1)
PE(1) MSG(2)
PE(1) MSG(3)
PE(1) MSG(4)
PE(1) MSG(5)
PE(2) MSG(0)
PE(2) MSG(1)
PE(2) MSG(2)
PE(2) MSG(3)
PE(2) MSG(4)
PE(2) MSG(5)
PE(3) MSG(0)
PE(3) MSG(1)
PE(3) MSG(2)
PE(3) MSG(3)
PE(3) MSG(4)
PE(3) MSG(5)
PE(4) MSG(0)
PE(4) MSG(1)
PE(4) MSG(2)
PE(4) MSG(3)
PE(4) MSG(4)
PE(4) MSG(5)
PE(5) MSG(0)
PE(5) MSG(1)
PE(5) MSG(2)
PE(5) MSG(3)
PE(5) MSG(4)
PE11(0)
PE11(1)
PE11(2)
PE11(3)
PE11(4)
PE11(5)
STRIPING
FACTOR
" C "
NPB

FIG. 3B

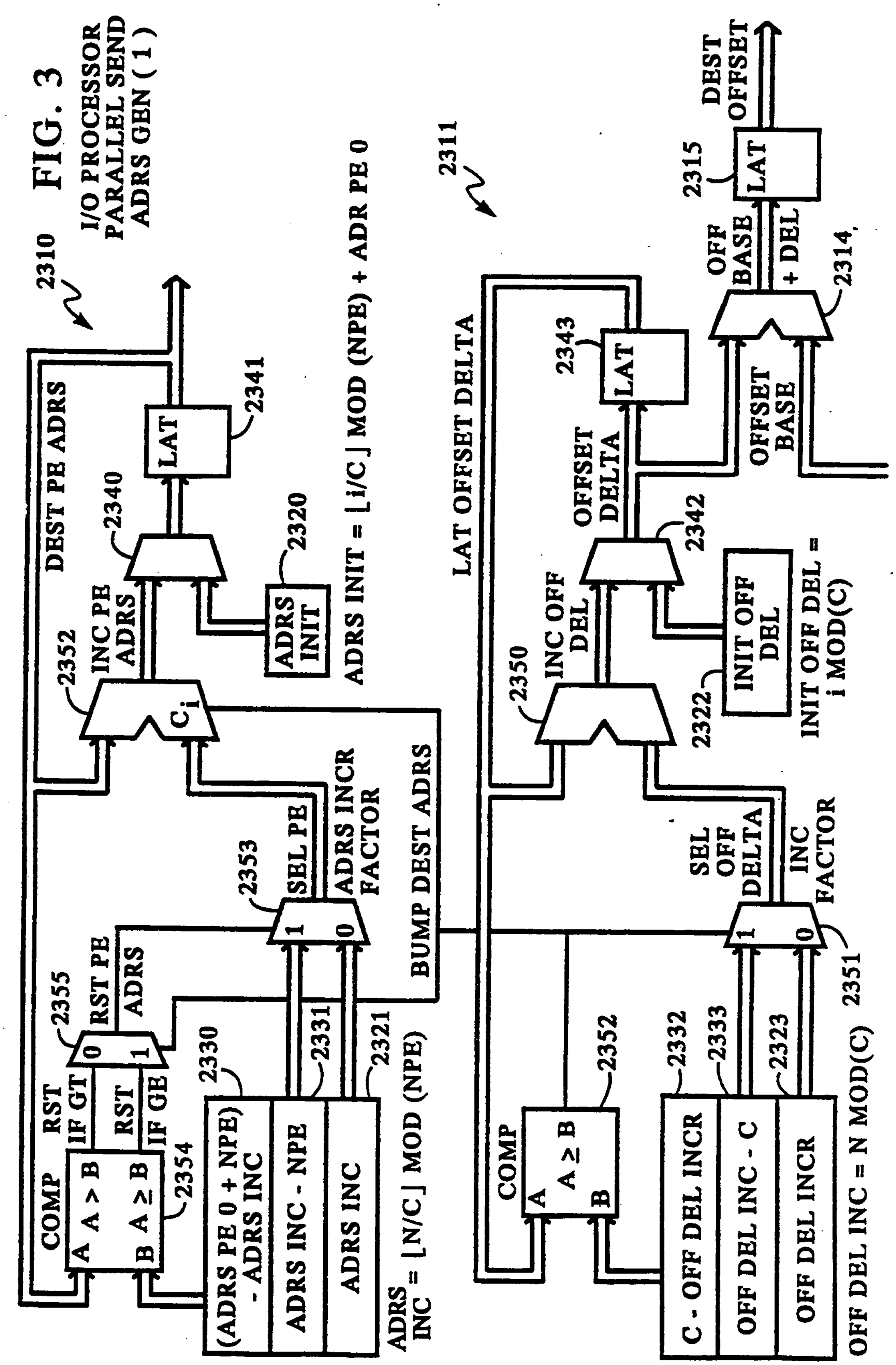

I/O PROCESSOR
PARALLEL SEND
ADRS GEN (1)
2310
DEST PE ADRS
LAT
2341
2340
INC PE ADRS
2352
C_i
ADRS INIT
2320
ADRS INIT = ⌊i/C⌋ MOD (NPE) + ADR PE 0
2353
SEL PE
ADRS INCR FACTOR
BUMP DEST ADRS
2355
RST PE ADRS
COMP
A > B
A ≥ B
RST IF GT
RST IF GE
2354
(ADRS PE 0 + NPE) - ADRS INC
ADRS INC - NPE
ADRS INC
2330
2331
2321
ADRS INC = ⌊N/C⌋ MOD (NPE)
2311
LAT OFFSET DELTA
2343
LAT
2315
DEST OFFSET
OFF BASE
+ DEL
2314
OFFSET BASE
OFFSET DELTA
2342
INC OFF DEL
2350
INIT OFF DEL
2322
INIT OFF DEL = i MOD(C)
SEL OFF DELTA
INC FACTOR
2351
COMP
A ≥ B
2352
C - OFF DEL INCR
OFF DEL INC - C
OFF DEL INCR
2332
2333
2323
OFF DEL INC = N MOD(C)

FIG. 3

FIG. 4A

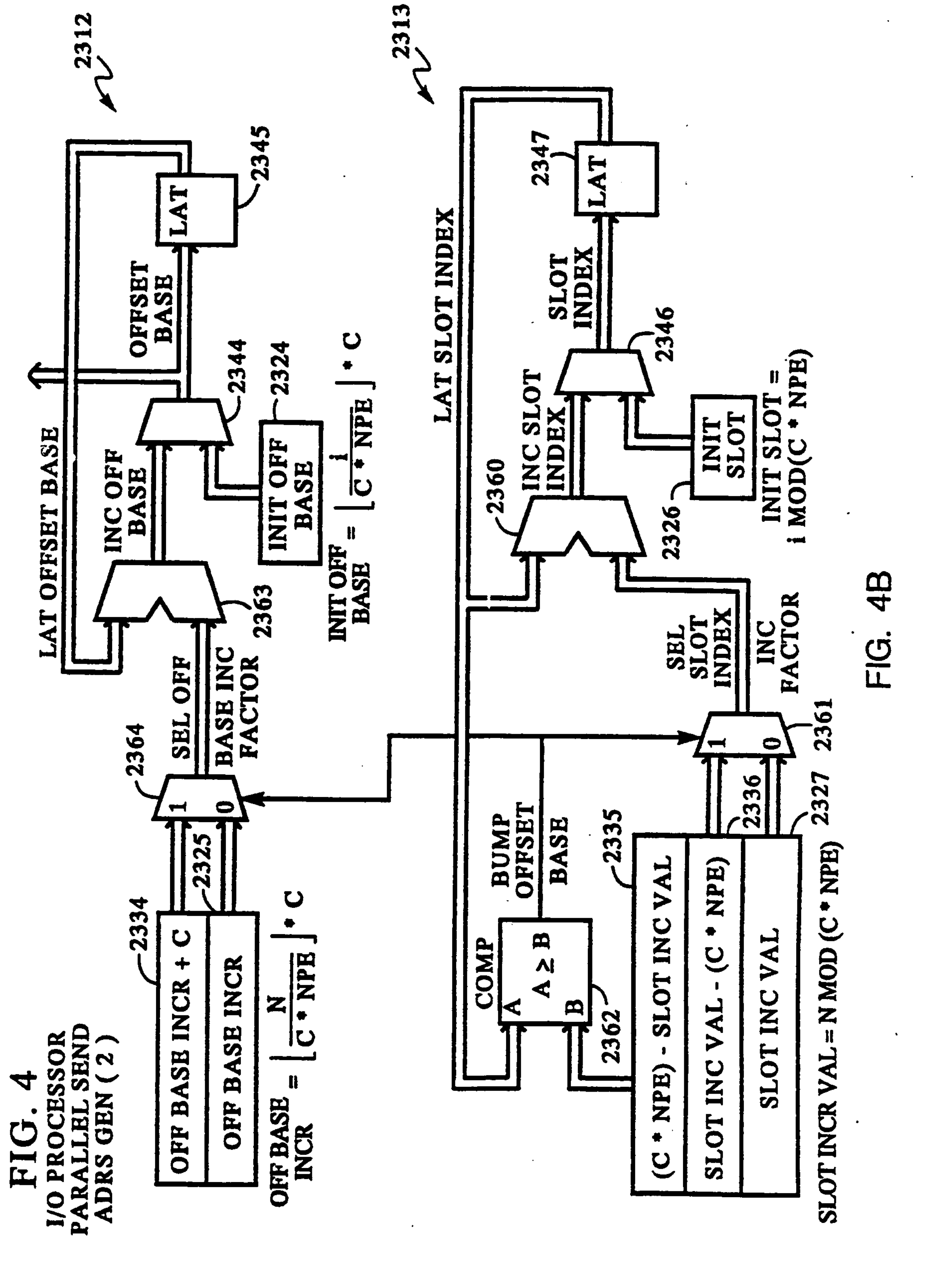


FIG. 4
I/O PROCESSOR PARALLEL SEND ADRS GEN (2)

FIG. 4B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/06848

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :G06F 15/80
US CL :395/275
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 395/275, 395/325,800,250

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Data,Network,Router,Routing,SPMD,SIMD,Leaf, Control,MIMD APS Searched

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US,A, 4,873,626 (Gifford) 10 October 1989 See Col. 4, lines 65 to Column 6, line 39., Column 13, lines 10-65, and Abstract.	1-39
Y	US,A, 4,598,400 (Hillis) 01 July 1986 See Abstract, and Col. 2, lines 30 to Column 3, line 25.	1-39
A	US,A, 5,111,389 (McAuliffe et al.) 05 May 1992 See Abstract and Col. 5, line 35 to Column 6, line 33.	
A	US,A, 4,910,669 (Gorin et al.) 20 March 1990 See Abstract.	

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 NOVEMBER 1992	07 DEC 1992
Name and mailing address of the ISA/ Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer ALPESH M. SHAH
Facsimile No. NOT APPLICABLE	Telephone No. (703) 308-0110